

A Power Gating Switch Box Architecture in Routing Network of SRAM-Based FPGAs in Dark Silicon Era

Zeinab Seifoori, Behnam Khaleghi, and Hossein Asadi

Data Storage, Networks, and Processing (DSN) Lab, Department of Computer Engineering
Sharif University of Technology, Tehran, Iran

Abstract—Continuous down scaling of CMOS technology in recent years has resulted in exponential increase in static power consumption which acts as a power wall for further transistor integration. One promising approach to throttle the substantial static power of *Field-Programmable Gate Array* (FPGAs) is to power off unused routing resources such as switch boxes, known as dark silicon. In this paper, we present a *Power gating Switch Box Architecture* (PESA) for routing network of SRAM-based FPGAs to overcome the obstacle for further device integration. In the proposed architecture, by exploring various patterns of used multiplexers in switch boxes, we employ a configurable controller to turn off unused resources in the routing network. Our study shows that due to the significant percentage of unused switches in the routing network, PESA is able to considerably improve power efficiency in SRAM-based FPGAs. Experimental results carried out on different benchmarks using VPR toolset show that PESA decreases power consumption of the routing network up to 75% as compared to the conventional architectures while preserving the performance intact.

I. INTRODUCTION

Despite major advantages of *Field-Programmable Gate Arrays* (FPGAs) such as shorter time-to-market, reduced *Non-Recurring Engineering* (NRE) cost, and design flexibility as compared to *Application-Specific Integrated Circuits* (ASICs), they still suffer from high power consumption. Several studies have revealed that the power consumption of FPGA-based designs is at least 7 to 14 times higher than that of ASICs [1]–[3]. One major contributor of the FPGA-ASIC power gap is static power which is attributed to continuous power consumption of transistors independent of switching frequency. Due to reducing the technology size in recent years and failure of *Denard Scaling* which circumscribes the clock rate [4], increasing rate of static power consumption is considerably higher than dynamic power which acts as a *power wall* against further scaling [5].

Two major sources of static power in FPGAs are logic blocks and routing resources. Previous studies show that utilization rate of interconnects, compared with logic blocks, is significantly low. On the other hand, it has been shown that a considerable part (greater than 70%) of static power consumption is consumed by interconnect resources [6], [7]. Therefore, the low utilization rate of interconnect resources in one hand, and substantial contribution of unused interconnect resources in the total power on the other hand necessitates an approach to alleviate the power consumption of interconnect resources. The building blocks of interconnect resources are programmable *Switch Boxes* (SBs) which are controlled by SRAM cells. Different arrangement of SBs results in different *Switch Matrix* (SM) topologies (i.e., *Subset*, *Wilton*, and *Universal*) with different routability¹ and topologies² [8]–[10]. The attributes (e.g., delay, area, and power) of a design mapped on an FPGA vary based on SM topology and SB structure.

Previous work aiming at reducing the routing static power can be classified into three categories. The first category attempts to cope with static power by exploiting non-power gating

techniques. Device-level low-power techniques such as *dual-vdd* and *dual-threshold* [11]–[13] fall into this category. Such low-power approaches, however, are not cost-efficient and increase the fabrication complexity. The second category has proposed novel logic architectures augmented with power-gating in which each logic block consists of a set of power efficient logic cells wherein only one cell is turned on based on the implemented function [14], [15]. The third category has targeted mainly to minimize static power through employing power gating in coarse- and fine-grained granularities which is applied in logic and/or routing resources of FPGA [16]–[23]. Most of the studies of this category apply dynamic (i.e., runtime) coarse-grained power gating technique on the modules that temporarily go idle. Such techniques, however, suffer from less power gating opportunities. Moreover, there are main challenges in applying power gating dynamically such as mitigating *In-Rush* current (i.e., the large current drawn from power lines by turning on the large power gated modules simultaneously), routing the power control signals (which are needed to turn on/off the power gated units dynamically), and modification in the CAD algorithms to provide more opportunities for power gating.

Considering the challenges of dynamic power gating techniques, a promising approach to improve the efficiency of static power consumption is to employ static power gating. Designing an efficient architecture that employs static power gating technique, however, requires comprehensive profiling of the device resources and their utilization patterns to carefully adapt the granularity. Inappropriate selection of granularity can negatively affect the power saving. Finer granularity power gating provides more opportunities for power saving but such opportunities come at the expense of more peripheral area and power overhead. Due to array-like topology of logic blocks, investigating the effect of granularity in logic blocks is straightforward while determining the most efficient granularity in routing network necessitates the examination of various SM topologies and SB structures which have not been done in the previous work.

In this paper, we investigate the routing resources to find an appropriate power gating granularity for different SM topologies and SB structures. In this regard, the main building block of SBs (i.e., multiplexers) are examined to extract the opportunity to turn off unused resources. Low utilization rate of multiplexers indicates that coarser level of power gating (e.g., SB level or SM level) is probably an effective approach to increase the power efficiency. To attain the best granularity, different granularities for power gating are proposed and their power consumption are estimated. Our results show that the most power efficient granularity in FPGA with specific SM topology and SB structure is not necessarily the best choice for other cases; thereupon, the optimum granularity should be chosen considering the SM topology and SB structure. In the proposed architecture, namely PESA, a power gating transistor is added to power supply of all SRAM cells and the corresponding multiplexer and buffers. An SRAM cell, namely PG-SRAM, is added to each power gating region to control its *on/off* state. PESA is scalable with different

¹Fewer required tracks per channel means higher routability.

²Determines which outgoing tracks can be connected to each incoming track.

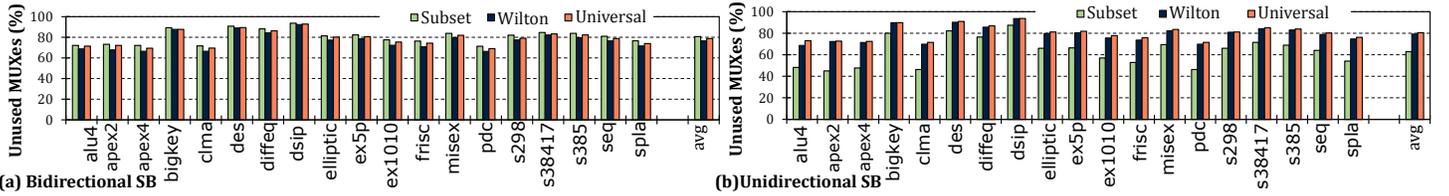


Fig. 1. Percentage of unused MUXes

granularities and appropriate granularity is chosen pursuant to SM topology and SB type.

We have evaluated our proposed architecture using VPR 7.0 toolset [24] in terms of static power consumption, area, and delay overhead. Different benchmark suites, e.g., MCNC and IWLS, have been used to show the efficiency of PESA over the baseline (i.e., SRAM-based FPGA with no power gating scheme). Experimental results show that PESA reduces the routing static power consumption, on average, up to 75% in specific topology. The area overhead ranges from 6% to 30% among different topologies. The results revealed that different SMs and SBs require different power gating architecture to provide optimum power gating solution.

Specifically, our novel contributions in this paper are:

- (1) We first analyze the routing resource utilization rates in different levels from the finest level to the coarsest one (i.e., SB multiplexers, SBs, and SMs) in detail.
- (2) Full examination of the effect of SM topologies and SB structures on different granularities of power gating is performed.
- (3) The most efficient granularity for each SM topology and SB structure with respect to area and delay overhead is proposed.
- (4) The impact of comprehensive industrial and standard benchmarks such as MCNC and IWLS on power gating granularity is evaluated and the efficiency of the proposed architecture is examined by aforementioned benchmark suites.

The rest of this paper is organized as follow. Section II represents the proposed architecture. Experimental setup and results are detailed in Section III, and finally, we conclude the paper in Section IV.

II. PROPOSED METHOD

The main objective of the proposed architecture is turning off inactive routing resources through power gating, which is a generally accepted approach to save the static power in both ASIC and FPGAs [25]–[28]. Although power gating is apparently a straightforward approach, its efficient implementation in FPGAs is very challenging and depends on the scattering form of utilized resources. If the overall resource usage rate is low, but the used resources are sporadic and uniformly scattered in the whole device, exploiting power gating will face with serious problems.

Consequently, to assess the efficiency of power gating in reduction of the static power consumption of routing network and finding the optimal power gating granularity in FPGAs, comprehensive information about utilization rate of interconnect resources (SBs) is needed. Due to the best area-delay trade-off of multiplexer-based switches [29] as well as their usage in commercial FPGAs [30], we focus our effort on multiplexer-based switches. Accordingly, in this section, we first analyze the overall utilization rate of interconnect resources (the multiplexer parts) and then we study different power gating granularity for SBs and estimate the power consumption of different SRAM-based cell structures with various granularities. Afterwards, the dependency of power consumption of proposed architectures to SM topology and SB structure is assessed. Lastly, the utilization pattern of multiplexers within SB (i.e., SB patterns) is examined to find the best granularity.

A. Resource Utilization

The interconnect of state-of-the-art FPGAs is composed of multiplexers, their selection bits, and corresponding output buffers that drive the output wire as a channel track. Approximately 60% of static power consumption is consumed by interconnect multiplexers [6]. In this section we target unused multiplexers, i.e., those with undriven/unconnected *output* (it should be noticed that some inputs of an unused multiplexer may be driven accidentally). We investigate the utilization rate of unused multiplexers in FPGAs with diverse SM topologies and SB structures by VPR 7.0 toolset over MCNC benchmarks. Fig. 1 represents the percentage of unused multiplexers in FPGAs with bidirectional (left) and unidirectional (right) SBs for different topologies of SMs (*Subset* [8], *Wilton* [9], and *Universal* [10]). According to this figure, the percentage of unused multiplexers is, on average, 79% and 74% for FPGAs with bidirectional and unidirectional SBs, respectively. The minimum ratio of unused multiplexers in unidirectional *Subset* SMs arises from more effective routability of this architecture that uses smaller channel width (thereby less unused resources).

Due to high unutilization rate of multiplexers, as shown in Fig. 1, it is concluded that if the overheads caused by power gating (i.e., area and delay) is at an acceptable level, multiplexer-level granularity can be very effective. Nevertheless, such high unutilization rate, particularly in bidirectional *Subset* topology suggests investigating coarser level of power gating to alleviate power and area overheads, because multiplexer-level granularity imposes nearly 25% area overhead (details are provided in Section III) which is considerable especially in unidirectional *Subset* topology with higher utilization rate of multiplexers.

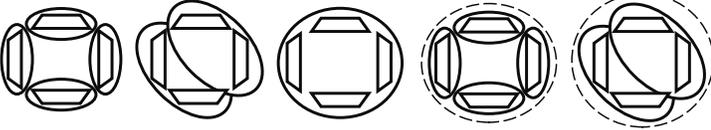
B. Granularity Assessment

According to high unutilization rate of multiplexers, different power gating granularities can be employed, including SM level (i.e., one power gating controller for an entire SM), SB level (i.e., one controller for each SB as shown in Fig. 2(c)), and intra-SB level. Power gating schemes for a SB are illustrated in Fig. 2. Each circle shows a power gating region associated with a controller (SRAM and switch). As demonstrated by Fig. 2(a), each of four multiplexers of a SB has a unique power gating controller. Fig. 2(d) represents a two-level power gating which acts similar to Fig. 2(a) but when all multiplexers are *off*, the whole regions are also power gated; hence, the four added controller SRAMs are also power gated. Approximately similar *off* and *on* state of adjacent multiplexers motivates us to propose the power gating structure shown in Fig. 2(b). As illustrated in this figure, each pair of multiplexers share one power gating controller. Fig. 2(e) shows two-level power gating corresponding to Fig. 2(b), analogous to the proposed structure of Fig. 2(d). This structure provides combination of coarse- and fine-grained power gating.

The power consumption of different SRAM-based cell structures (Fig. 2) can be estimated as reported in Table I. In this table, the power consumption of different structures is expressed in term of α (unutilization probability of each multiplexer), P_S (the power consumption of one SRAM cell), and P_M (the power consumption of each multiplexer including the power consumption of mul-

TABLE I. POWER CONSUMPTION OF DIFFERENT SRAM-BASED CELL STRUCTURES

SRAM-based cell structures	Power consumption
(a) $Arch_{SB,4}$	$P_{SB,4} = \sum_{i=0}^4 \binom{4}{i} \alpha^i (1-\alpha)^{4-i} (P_M \cdot i \cdot k + P_M(4-i)) + 4P_S$
(b) $Arch_{SB,2}$	$P_{SB,2} = \alpha^4 (4kP_M) + 2\alpha^2 (1-\alpha)^2 (2kP_M + 2P_M) + (1-\alpha^4 - 2\alpha^2(1-\alpha)^2)(4P_M) + 2P_S$
(c) $Arch_{SB,1}$	$P_{SB,1} = \alpha^4 (4P_M \cdot k) + (1-\alpha^4)(4P_M) + P_S = \alpha^4 (4kP_M) + (1-\alpha^4)(4P_M) + P_S$
(d) $Arch_{SB,4,1}$	$P_{SB,4,1} = \alpha^4 (4P_M + 4P_S) \cdot k + \sum_{i=0}^3 \binom{4}{i} \alpha^i (1-\alpha)^{4-i} (P_M \cdot i \cdot k + P_M(4-i) + 4P_S) + P_S$
(e) $Arch_{SB,2,1}$	$P_{SB,2,1} = \alpha^4 ((4P_M + 2P_S) \cdot k) + 2\alpha^2 (1-\alpha)^2 (2kP_M + 2P_M + 2P_S) + (1-\alpha^4 - 2\alpha^2(1-\alpha)^2)(4P_M + 2P_S) + P_S$



(a) Arch (SB,4) (b) Arch (SB,2) (c) Arch (SB,1) (d) Arch (SB,4,1) (e) Arch (SB,2,1)
Fig. 2. Different granularities for SB power gating

tiplexer, buffer, and two SRAM cells used for selection lines of multiplexer). We assume that power consumption of multiplexer at *off* state is K times of its *on*-state ($P_{M,off} = K \times P_{M,on}$). Based on our experiments which will be detailed in Section III, the factor K is about 0.1, on average. For instance, the first row of Table I corresponds to the power consumption of SB structure represented in Fig. 2(a). According to the number of used multiplexers per SB, five different conditions can occur in a SB with different possibilities. The probability of having i unused multiplexer is $\binom{4}{i} \times \alpha^i \times (1-\alpha)^{4-i}$, which will consume a power equal to $i \times P_M \times K$ (each of i multiplexers consumes P_M multiplied in power gated factor K), and the rest $4-i$ used multiplexers consume $(4-i) \times P_M$. Finally, the power overhead of four control SRAMs is included in the estimated power. Fig. 3 illustrates the normalized power consumption of different SRAM-based cell structures against the baseline FPGA architecture which is a standard device without power gating. The horizontal axis is the unutilization probability of each multiplexer (α), which ranges from 0 (indicates that all multiplexers are used) to 1 (indicates that all multiplexers are unused) and the vertical axis is the estimated power consumption. These curves are obtained based on formulas presented in Table I where P_M and P_S parameters are obtained through circuit-level simulation using HSPICE. According to the experiments, P_M is 7 times of P_S and hence the power consumption of a SB is approximately 28 times of P_S . As it is obvious from Fig. 3, for the range of obtained unutilization rate (60%-80%), SB,4 architecture (shown in Fig. 2(a)) and SB,4,1 architecture (shown in Fig. 2(d)) provide the optimum power efficiency.

When estimating the power consumption of various SB structures (Table I), it is assumed that utilized routing resources are distributed on the device uniformly and the utilization probability of a multiplexer is independent of the utilization probability of adjacent multiplexers. However, the distribution of routing resources are not uniform in the mapped designs, (e.g., corner SMs are rarely used while the regions beside high-fanout nets are congested) and the utilization probabilities of multiplexers are tied together, as well. As a result, comprehensive information about distribution of used multiplexers is needed for an accurate calculation of power consumption experiments and this information necessitates conducting experimental analysis.

C. Topology Dependence

Referring to Fig. 1, only a quarter of total multiplexers in routing resources is used and the others are unused. For finding the best granularity, the scattering model of used multiplexers should be investigated. If the used multiplexers are congested at SMs, i.e., the majority of SBs within a used SM are used, and furthermore, high percentage of SB multiplexers are utilized (which means that used multiplexers are congested in particular

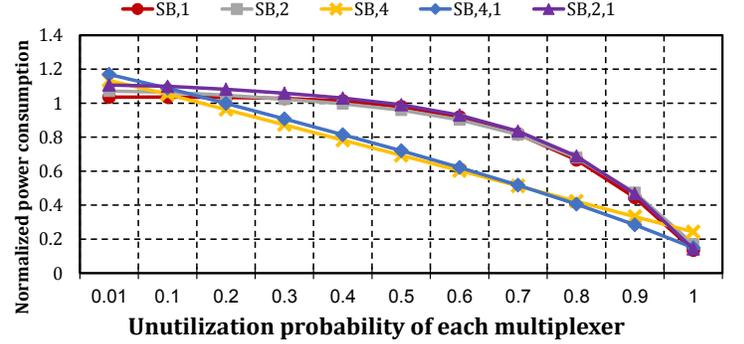


Fig. 3. Normalized power consumption of different SRAM-based cell structures against the baseline

SMs), then power gating at the SM level would be effective. As the second level of granularity, if the used multiplexers are congested in SBs, i.e., the SBs are either fully-used or unused, then power-gating at the SB level would be more promising. Finally, if the used multiplexers are sporadic, majority of SBs and thereby SMs are (partly) used and applying power gating in SM and SB granularity is not useful and more fine grained power gating should be exploited.

The utilization probability of a SM with bidirectional and unidirectional structures can be estimated as follow:

$$\alpha_{SM} = \begin{cases} \alpha^{4 \times \frac{W}{2}} & \text{Unidirectional} \\ \alpha^{4 \times W} & \text{Bidirectional} \end{cases}$$

In this equation, α_{SM} indicates the unutilization probability of SM and W stands for channel width. α^4 is the probability of unused SB (all four multiplexers within the SB are unused). Note that there are $W/2$ and W SBs in uni- and bidirectional SB, respectively. As it will be discussed later, due to the fact that benchmarks act completely different from the analysis, analysis alone is insufficient to evaluate the resource utilization of the device and benchmarks should be examined thoroughly. For example, the unutilization probability of multiplexers (α) for *alu4* benchmark, as reported in Fig. 1, with bidirectional and unidirectional SBs is 72% and 48%, respectively. Therefore, the unutilization probability of SMs (α_{SM}) for this benchmark with channel width 42 and 44 for bidirectional and unidirectional SBs, respectively, is about 5.5×10^{-23} and 8.9×10^{-29} . However, the achieved unutilization probability of SMs for bidirectional and unidirectional SBs through benchmark examination is about 4.2% and 4.1%, as it has been demonstrated in Fig. 4. As the same way, the unutilization probability of SMs for MCNC benchmarks obtained by benchmark examination is, on average, 2.07% and 2.33% for bidirectional and unidirectional SBs, respectively (see Fig. 4). Nevertheless, the analytically achieved probability of SMs for bidirectional and unidirectional SBs is on average 7.33×10^{-28} and 1.27×10^{-30} , respectively. As it stands, the achieved results from benchmarks examination are quite different from the results presented by this analysis and, hence, it does not suffice to rely on only analysis in order to extract the resource utilization.

As another example, the unutilization probability of SBs (α^4) using the unutilization probability of multiplexers (reported in Fig. 1) can be estimated as 40%, 33%, and 37% for FPGA with bidirectional *Subset*, *Wilton*, and *Universal* SBs. As it is shown in Fig. 5, the unutilization probability of SBs obtained through experiments for MCNC benchmark is on average 45%, 85%,

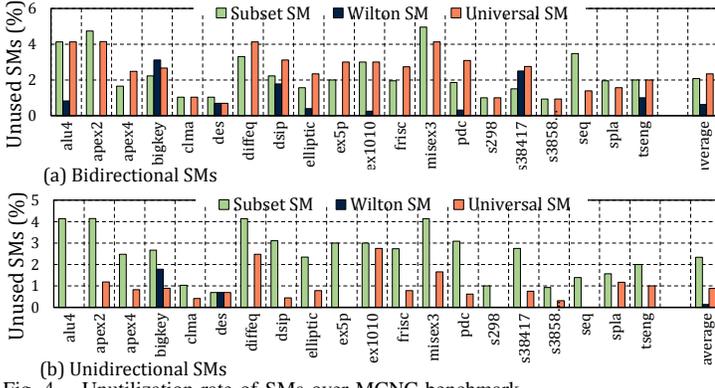


Fig. 4. Unutilization rate of SMs over MCNC benchmark

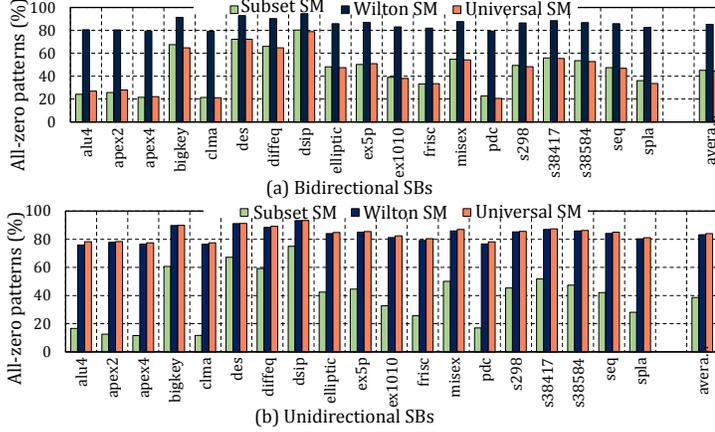


Fig. 5. Repetition rate of all zero pattern for FPGA with various topologies of SB over MCNC benchmark

and 44% for the same architectures, respectively. This reveals that while the analytically obtained utilization probabilities are roughly similar, the actual unutilization rates significantly vary according to SM topology.

Comparison between the obtained results through experiments (reported in Fig. 5) reveals that the unutilization probability of SBs in MCNC benchmarks for FPGAs with uni- and bidirectional *Wilton* SBs is much higher than FPGAs with uni- and bidirectional *Subset* SBs. This observation confirms the significant effect of SB topology on the unutilization probability of routing resources. On the other hand, the difference between unutilization probability of SBs in MCNC benchmarks for FPGA with uni- and bidirectional *Universal* SBs (as shown in Fig. 5) demonstrates the impact of SM structure on unutilization probability of routing resources. Likewise, *alu4* and *clma* benchmarks (from MCNC benchmarks) on FPGA with bidirectional *Subset* SB with approximately the same α (unutilization probability of multiplexers), as reported in Fig. 1, have different probability of unutilization SMs (see Fig. 4). Therefore, this indicates the effect of benchmarks on the unutilization probability of routing resources.

D. SB Configuration Dependence

Although conducted studies on investigating the ratio of unutilized routing resources such as SMs and SBs are necessary, the analysis method by itself is not suitable, because it cannot properly characterize the resource utilization. The similar concept is also valid for partly-used SBs which should be taken into account in proposing fine-grained intra-SB power gating regions. For example, considering the case of SMs with *Wilton* topology, where 80% of its SBs are unused, the power consumption of the most coarse-grained architecture (Fig. 2(c)) is obtained through the following equation ($P_M = 7P_S$):

$$P_{SB,4} = 0.8 \times 4KP_M + 0.2 \times 4P_M + P_S = 8.8P_S$$

Whereas the power consumption of the most fine-grained architecture (Fig. 2(a)) can be estimated as follows:

$$P_{SB,1} = 0.8 \times 4KP_M + 0.2 \times (i \times P_M + (4-i) \times KP_M) + 4P_S$$

In this equation, i denotes the number of used multiplexers in SB. The power consumption of the SB,4 architecture in the best scenario, which three out of four multiplexers in underutilized SBs are unused (i.e., $i=1$), is about $8.1P_S$ and is lower than the power consumption of the most coarse-grained architecture. Meanwhile, the power consumption of this architecture, when one out of four multiplexers in underutilized SBs is unused (i.e., $i=3$), is about $10.6P_S$ which is higher than the power consumption of the most coarse-grained architecture.

Thereupon, the power consumption of the proposed architectures significantly depends on the number of utilized multiplexers within used SBs. Assuming two out of four multiplexers in a used SBs are utilized, two scenarios for multiplexers within SBs can happen; whether two used multiplexers belong to the same power gating group or not. The total power consumption of SB,2 architecture for these two scenarios can be obtained by Equation (1) and Equation (2). Equation (1) represents the SB power consumption when the used multiplexers within SB belong to the same power gating group and Equation (2) represents the SB power consumption when the used multiplexers within SB do not belong to the same power gating group (hence, cannot be turned off).

$$P_{SB,2} = 0.8 \times 4KP_M + 0.2(2P_M + 2KP_M) + 2P_S = 7.3P_S \quad (1)$$

$$P_{SB,2} = 0.8 \times 4KP_M + 0.2(4P_M) + 2P_S = 9.8P_S \quad (2)$$

Consequently, the power consumption of the proposed architecture is also dependent on the pattern of utilized multiplexers within SBs. As the results indicate, to accurately estimate the power consumption of different power gating granularity, in addition to the utilization rate of SBs, the SB configuration should also be examined.

Taken together, if the unutilization rate of SBs is high, according to the utilization rate of multiplexers within the used SBs, the most fine-grained architecture can either aggravate or enhance the power consumption as compared to the most coarse-grained architecture. Furthermore, the power consumption of SB,2 architecture can change due to the possible arrangement of utilized multiplexer within used SBs.

SB patterns: Since each SB comprises four multiplexers, we consider a SB pattern as a 4-bit sequence which indicates the used/unused state of multiplexers. For considered 4-bit SB pattern (i.e., $b_1b_2b_3b_4$), b_1 to b_4 correspond to multiplexers on side 1 to 4, respectively. The select bits of unused multiplexers are “00” and the state of this multiplexer in 4-bit SB pattern is “0”. Otherwise, the multiplexer is utilized and its state in 4-bit SB pattern is “1”. In order to provide more opportunities for power saving in SBs, we have extracted the frequency of different SB patterns to find out if there is some particular patterns that have higher repetition rate than others. Fig. 6(a) and Fig. 6(b) show the most frequent SB patterns and their repetition rates for variety of topologies of bidirectional and unidirectional SB over MCNC benchmarks, respectively. As it can be observed, in most frequent patterns, unused multiplexers are next to each other (denoted by consecutive zeros) and their power consumption can be controlled as a group. According to the patterns, an efficient approach is utilizing the power gating in finer granularity and grouping the multiplexers and their corresponding configuration bits into few sets. Given that the achieved results in this section for different types of SBs may be different, the best granularity from the aspect of power efficiency and area overhead will be discussed in Section III.

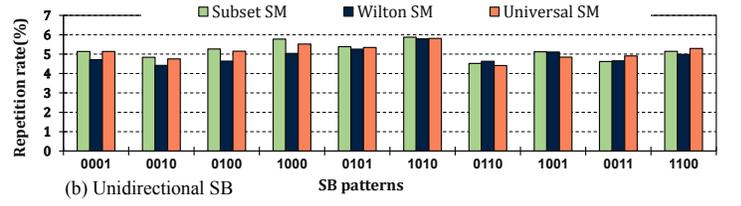
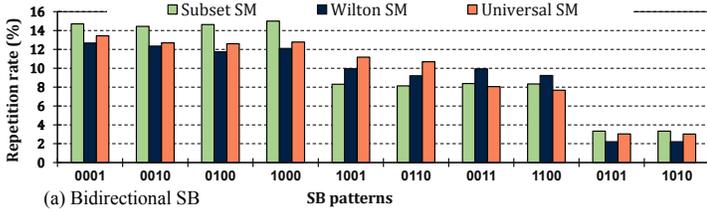


Fig. 6. Most frequent patterns for FPGA with various topologies of SB over MCNC benchmark

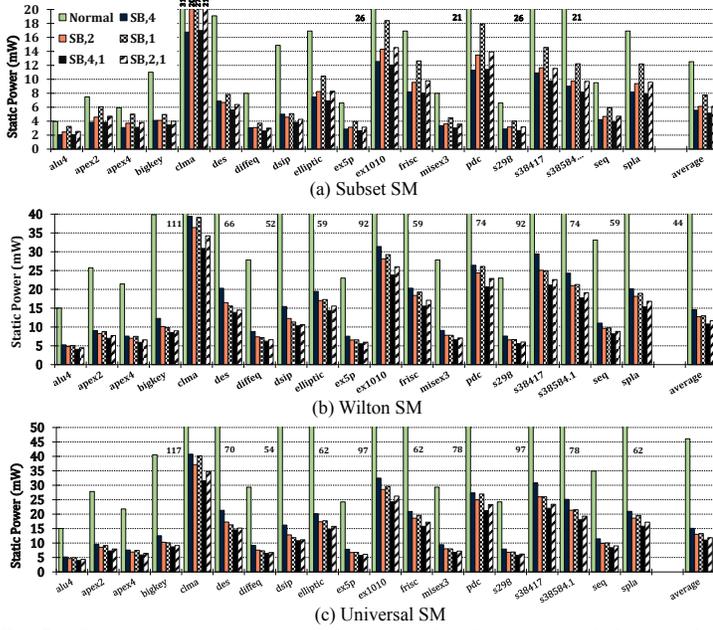


Fig. 7. Static power consumption in the proposed architectures for bidirectional *Subset*, *Wilton*, and *Universal* SMs over MCNC benchmarks

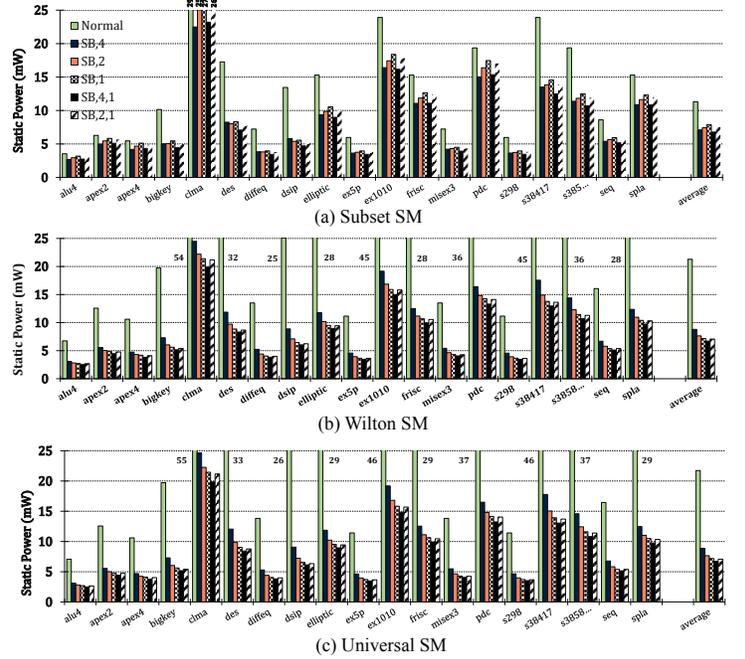


Fig. 8. Static power consumption in the proposed architectures for unidirectional *Subset*, *Wilton*, and *Universal* SMs over MCNC benchmarks

III. EXPERIMENTAL SETUP AND RESULTS

In this section, we detail the experimental setup and the results of power saving achieved by the proposed architecture. In order to evaluate the proposed architecture, we implement different benchmark circuits including MCNC and IWLS benchmarks using VPR. Power consumption of different architectures are measured by HSPICE circuit-level simulations using *Predictive Technology Model* (PTM) [31]. The typical minimum-size six-transistor SRAM cells, which are employed in FPGAs, are used and the size of transistors and buffers are obtained from VTR repository. The acquired transistor size is $1.8X$ the minimum width ($1.8 \times 90nm$) and the size of considered buffer is $5X$ which is sufficient to drive the wires with segment length of $L = 1$. Finally, we assume an FPGA with 6-input LUTs and 10 LUTs per logic block and SMs with $F_s = 3$. Overhead of each architecture includes the area of PG-SRAM(s) and the cut-off transistor with $W = 5W_{min}$ which is reported in Table II.

The power consumption of traditional FPGAs with *Subset* switch type is compared with five proposed architectures for MCNC benchmarks in Fig. 7(a). As it is shown in this figure, the best power efficiency is achieved by the architectures with five PG-SRAM per SB (Fig. 2(d)). Hence, since the saved power attained by *SB,4* architecture and *SB,4,1* architecture are almost the same, for bidirectional *Subset*, the architecture with four PG-SRAM cells per SB due to its less area overhead is preferred. The *SB,4* architecture imposes 25%, while the *SB,4,1*

architecture imposes 31% area overhead to the routing fabric. The experimental results shown in Fig. 7(b) and Fig. 7(c) indicate that for FPGAs with bidirectional *Wilton* and *Universal* SBs, the *SB,4,1* and *SB,2,1* are the most efficient architectures among the proposed architectures but *SB,2,1* architecture, due to imposing less area overhead to routing fabric, is preferable. Comparing the power consumption of proposed architectures in FPGAs with different bidirectional topologies reveals that although the *SB,4* architecture in *Subset* SM belongs to the most power-efficient architectures, it consumes higher power in FPGAs with *Wilton* and *Universal* SMs than the other proposed architectures. In overall, the power consumption of *SB,4,1* and *SB,4* architectures for FPGAs with *Subset* SMs are minimum but their area overhead on routing fabric are the most. This is while power consumption of the *SB,4,1* and *SB,2,1* architectures in FPGAs with *Wilton* and *Universal* SBs are minimum (shown in Fig. 7 and Fig. 8).

Fig. 9 illustrates the power-area product of proposed architectures over MCNC benchmarks. Taking into account the power-area product, *SB,1* will be the most efficient architecture. Fig. 10 reports power saving achieved through power gating with different granularities over MCNC benchmarks as compared to *SB granularity*. As is shown in this figure, the proposed architectures improves the effectiveness of conventional power gating architecture up to 40%.

Fig 11 illustrates the normalized power saving obtained by the proposed architecture over IWLS benchmarks with respect to the baseline (i.e., SRAM-based FPGA with no power gating scheme). For the sake of brevity, detailed results of IWLS benchmarks are removed. As shown in Fig 11, the *SB,4,1* is the most power efficient architecture but imposes maximal area overhead, as well. The *SB,1* architecture affords the best power-area product.

TABLE II. AREA OVERHEAD OF PROPOSED ARCHITECTURE

Architecture	SB,4	SB,2	SB,1	SB,4,1	SB,2,1
Area overhead	25.1%	12.5%	6.2%	31.3	18.8%

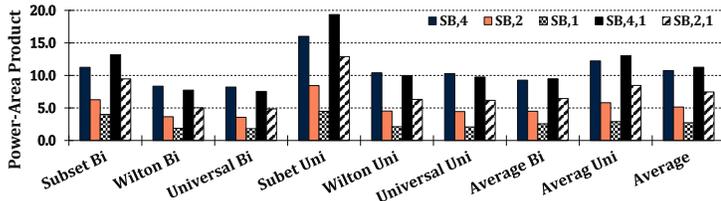


Fig. 9. Power-area product of proposed architectures for *Subset*, *Wilton*, and *Universal* SMs over MCNC benchmarks

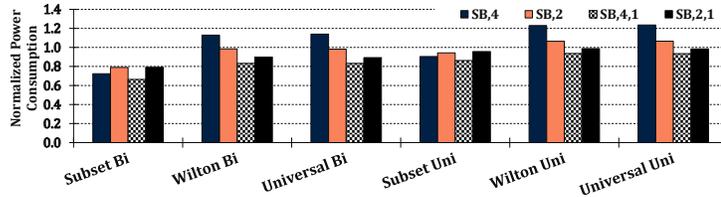


Fig. 10. Normalized power consumption of proposed architectures with respect to *SB,1* architecture for *Subset*, *Wilton*, and *Universal* SMs over MCNC benchmarks

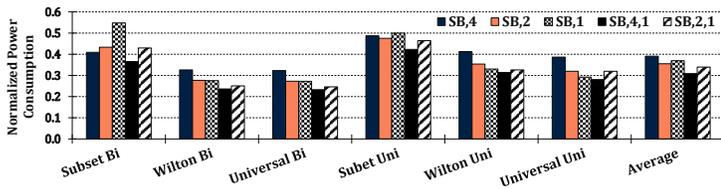


Fig. 11. Normalized power consumption of proposed architectures with respect to the baseline over IWLS benchmarks

IV. CONCLUSION

In this paper, we presented different power gating architectures to reduce the static power consumption in the routing network of SRAM-based FPGAs in the dark silicon era. As the experimental results demonstrated, our proposed architecture reduces the static power consumption up to 75%. Considering the architecture with the best power-area product, the power consumption is reduced by 57%. In addition, it is shown that the efficiency of a power gating architecture is highly correlated with the SM topology, SB structure, the implemented design, and the pattern of utilized multiplexers within SBs.

REFERENCES

- [1] I. Kuon and J. Rose, "Measuring the gap between FPGAs and ASICs," *IEEE Transactions on computer-aided design of integrated circuits and systems*, vol. 26, no. 2, pp. 203–215, 2007.
- [2] S. D. Brown, R. J. Francis, J. Rose, and Z. G. Vranesic, *Field-programmable gate arrays*. Springer Science & Business Media, 2012, vol. 180.
- [3] P. S. Zuchowski, C. B. Reynolds, R. J. Grupp, S. G. Davis, B. Cremen, and B. Troxel, "A hybrid ASIC and FPGA architecture," in *Proceedings of the 2002 IEEE/ACM international conference on Computer-aided design*. ACM, 2002, pp. 187–194.
- [4] R. H. Dennard, F. H. Gaensslen, L. Kuhn, and H. Yu, "Design of micron MOS switching devices," in *Electron Devices Meeting, 1972 International*, vol. 18. IEEE, 1972, pp. 168–170.
- [5] M. B. Taylor, "A landscape of the new dark silicon design regime," *IEEE Micro*, vol. 33, no. 5, pp. 8–19, 2013.
- [6] T. Tuan and B. Lai, "Leakage power analysis of a 90nm FPGA," in *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*. IEEE, 2003, pp. 57–60.
- [7] V. Degalahal and T. Tuan, "Methodology for high level estimation of FPGA power consumption," in *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*. ACM, 2005, pp. 657–660.
- [8] G. G. Lemiex and S. D. Brown, "A detailed routing for allocating wire segments in field-programmable gate arrays," in *ACM Physical Design Workshop, Lake Arrowhead, CA, 1993*, pp. 215–226.
- [9] S. J. Wilton, "Architectures and algorithms for field-programmable gate arrays with embedded memory," Ph.D. dissertation, Citeseer, 1997.
- [10] Y.-W. Chang, D. Wong, and C. Wong, "Universal switch modules for FPGA design," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 1, no. 1, pp. 80–101, 1996.

- [11] J. H. Anderson and F. N. Najm, "Low-power programmable FPGA routing circuitry," *IEEE transactions on very large scale integration (VLSI) systems*, vol. 17, no. 8, pp. 1048–1060, 2009.
- [12] R. Krishnan and J. P. de Gyvez, "Low energy switch block for FPGAs," in *VLSI Design, 2004. Proceedings. 17th International Conference on*. IEEE, 2004, pp. 209–214.
- [13] M. Klein, "WP298: Power Consumption at 40 and 45 nm," 2009.
- [14] A. Ahari, B. Khaleghi, Z. Ebrahimi, H. Asadi, and M. B. Tahoori, "Towards dark silicon era in fpgas using complementary hard logic design," in *2014 24th International Conference on Field Programmable Logic and Applications (FPL)*. IEEE, 2014, pp. 1–6.
- [15] Z. Ebrahimi, B. Khaleghi, and H. Asadi, "PEAF: A power-efficient architecture for SRAM-Based FPGAs using reconfigurable hard logic design in dark silicon era," *IEEE Transactions on Computers (TC), Special Section on Innovation in Reconfigurable Computing Fabrics: from Devices to Architectures*, 2017.
- [16] S. Yazdandshenas and H. Asadi, "Fine-grained architecture in dark silicon era for sram-based reconfigurable devices," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 10, pp. 798–802, 2014.
- [17] Y. Lin, F. Li, and L. He, "Routing track duplication with fine-grained power-gating for FPGA interconnect power reduction," in *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*. ACM, 2005, pp. 645–650.
- [18] A. A. Bsoul and S. J. Wilton, "An FPGA architecture supporting dynamically controlled power gating," in *Field-Programmable Technology (FPT), 2010 International Conference on*. IEEE, 2010, pp. 1–8.
- [19] A. Gayasen, Y. Tsai, N. Vijaykrishnan, M. Kandemir, M. J. Irwin, and T. Tuan, "Reducing leakage energy in FPGAs using region-constrained placement," in *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays*. ACM, 2004, pp. 51–58.
- [20] R. P. Bharadwaj, R. Konar, P. T. Balsara, and D. Bhatia, "Exploiting temporal idleness to reduce leakage power in programmable architectures," in *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*. ACM, 2005, pp. 651–656.
- [21] C. Li, Y. Dong, and T. Watanabe, "New power-aware placement for region-based FPGA architecture combined with dynamic power gating by PCHM," in *Proceedings of the 17th IEEE/ACM international symposium on Low-power electronics and design*. IEEE Press, 2011, pp. 223–228.
- [22] A. A. Bsoul and S. J. Wilton, "An FPGA with power-gated switch blocks," in *Field-Programmable Technology (FPT), 2012 International Conference on*. IEEE, 2012, pp. 87–94.
- [23] C. H. Hoo, Y. Ha, and A. Kumar, "A directional coarse-grained power gated FPGA switch box and power gating aware routing algorithm," in *2013 23rd International Conference on Field programmable Logic and Applications*. IEEE, 2013, pp. 1–4.
- [24] J. Luu, J. Goeders, M. Wainberg, A. Somerville, T. Yu, K. Nasartschuk, M. Nasr, S. Wang, T. Liu, N. Ahmed *et al.*, "VTR 7.0: Next generation architecture and CAD system for FPGAs," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, vol. 7, no. 2, p. 6, 2014.
- [25] F. Li, Y. Lin, L. He, and J. Cong, "Low-power FPGA using pre-defined dual-Vdd/dual-Vt fabrics," in *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays*. ACM, 2004, pp. 42–50.
- [26] B. H. Calhoun, F. A. Honore, and A. P. Chandrakasan, "A leakage reduction methodology for distributed MTCMOS," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 5, pp. 818–826, 2004.
- [27] T. Kuroda and T. Sakurai, *Low-voltage technologies*. Design of High Performance Microprocessor Circuits, IEEE Press., 2001.
- [28] T. Tuan, S. Kao, A. Rahman, S. Das, and S. Trimberger, "A 90nm low-power FPGA for battery-powered applications," in *Proceedings of the 2006 ACM/SIGDA 14th international symposium on Field programmable gate arrays*. ACM, 2006, pp. 3–11.
- [29] C. Chen, R. Parsa, N. Patil, S. Chong, K. Akarvardar, J. Provine, D. Lewis, J. Watt, R. T. Howe, H.-S. P. Wong *et al.*, "Efficient FPGAs using nanoelectromechanical relays," in *Proceedings of the 18th annual ACM/SIGDA international symposium on Field programmable gate arrays*. ACM, 2010, pp. 273–282.
- [30] D. Lewis, E. Ahmed, G. Baeckler, V. Betz, M. Bourgeault, D. Cashman, D. Galloway, M. Hutton, C. Lane, A. Lee *et al.*, "The Stratix II logic and routing architecture," in *Proceedings of the 2005 ACM/SIGDA 13th international symposium on Field-programmable gate arrays*. ACM, 2005, pp. 14–20.
- [31] (2013 (accessed March 20, 2016)) Predictive technology model (ptm). [Online]. Available: <http://ptm.asu.edu/>